数据引领新闻业的未来

摘　要：大数据时代的到来，为传统新闻业带来了新的机遇。数据新闻的诞生和发展，丰富了新闻报道的内容和方式。本文从数据新闻的定义出发，梳理了数据新闻的起源、发展和报道方式，并通过 ProPublica 和 The Upshot 这两个数据新闻平台的的多个案例，对数据新闻报道的阶段和其所呈现的特点做出思考与总结。

关键词：数据新闻；数据；个性化；可视化

中图分类号：G210　　**文献标识码**：A

文章编号：1671-0134（2018）08-112-03　　**DOI**：10.19483/j.cnki.11-4653/n.2018.08.046

文 / 曹慎慎

引言

数据和算法正在改变着各个行业，重新塑造我们的生活。现今，讲故事也要靠数据了。几年前，老牌新闻名校“哥伦比亚大学”开设了一个名为 Lede 的新闻学特色项目，一经推出便迅速成为数据新闻教育的示范性项目。

该项目针为计算机方面零基础或者基础薄弱，但是有志于学习计算机编程的学生提供了两个选择：其一，参加为期 12 周的加强型夏季训练营，授课内容有计算基础（Foundations of Computing）、数据和数据库（Data and Databases）、算法（Algorithms）和数据平台分析研究（Data Analysis Studio）等；其二，学生参加完 12 周的训练营之后，可以根据自己的兴趣自行选择是否继续秋季课程的学习。秋季课程将更加系统地学习数据分析和算法，课程也围绕基本数据结构（Essential Data Structures）、社会科学数据分析（Data Analysis for Social Sciences）、用数据讲故事（Storytelling with Data）、数字行动（Digital Activism）等。可以看出，Lede 项目的课程以社会科学和人文科学为基础，为了让学生们为将来从事以数据统计为驱动的媒体业务做好准备。

1. 数据生成新闻

尽管近年来人们对大数据的提法并不陌生，但对于什么是数据新闻（Data Journalism），大家的理解和探索还只是开始。乔纳森·格雷（Jonathan Gray）等人在《数据新闻手册》（The Data Journalism Handbook，2012）中这样写道，从表面上讲，数据新闻就是用数据报道新闻，但是“数据”和“新闻”本身就是让人困扰的术语，尤其当我们生活在一个方方面面都可以用数字描绘的时代，数据可以为新闻提供来源，也可以成为讲故事的工具，或者说数据与新闻互为因果，它是一种“将记者的新闻素养与运用大数据信息结合起来报道新闻的方式”；特洛伊·西伯德克斯（Troy Thibodeaux）在《五步开始数据新闻》（5 tips for getting started In data Journalism）中写道，“将数据运用到制作和传播信息的过程中，它反映了内容生产者和设计、计算机和统计领域之间的互动”；国内学者章戈浩认为，数据新闻是“对数据进行分析与过滤，从而创作出新闻报道的方式”；方洁等人则提出，数据新闻是“基于数据的抓取、挖掘、统计、分析和可视化呈现的新型新闻报道方式”。

西蒙·罗杰斯（Simon Rogers）是一位经验丰富的数据新闻编辑，他曾在《卫报》、推特和谷歌从事数据分析，同时，他也是数据新闻博客（Datablog）的创始人。2013 年，罗杰斯出版了《事实是神圣》（Facts are sacred）一书，回顾和梳理了他曾在《卫报》工作 15 年期间做出的数据新闻报道，讲述了数据新闻背后的故事，数据如何改变我们的生活以及我们能从中学到什么。在罗杰斯看来，数据新闻的首要任务是讲故事，而不是画图形或写代码。数据是开放的，属于每一个人，在数据的呈现下，人们可以更好地理解新闻。与此同时，随着数字技术的发展和更迭，总会有越来越多的工具帮助媒体进行数据分析和结果处理。

2. 数据新闻的报道方式

2010 年 8 月，德国之声记者米尔科·洛伦兹（Mirko Lorenz）提出了进行数据新闻报道的四个步骤：首先是数据挖掘；其次是数据过滤；再次是，数据可视化；最后是制作新闻。与之类似的是罗杰斯在其个人博客中对“数据新闻流程”所做的介绍。罗杰斯认为，数据新闻呈现的是一个多线性、全方位的报道流程，一方面处理数据，另一方面不断验证、质询数据的信度与价值，最后通过

多种手段与渠道发布报道。

伯明翰城市大学教授保尔·布拉德肖（Paul bradshaw）在《数据新闻的倒金字塔结构》中提出了“双金字塔模型”，更形象地揭示了数据在传播过程中的变化。其中，对于数据处理的过程，布拉德肖用倒金字塔结构来呈现，分别是数据汇编（Compile）、数据整理（Clean）、了解数据（Context）和数据整合（Combine）四部分。数据处理的目的是实现有效传播，而数据新闻传播则以“正金字塔结构”进行，布拉德肖认为应当在报道中体现出可视化（Visualise）、叙事化（Narrate）、社会化（Socialise）、人性化（Human）。

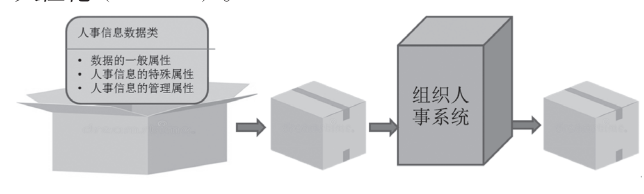


图 1　数据新闻“双金字塔模型”

事实上，不论是以何种方式进行数据新闻报道，数据获取、处理和呈现是其中不可缺少的三个步骤。谈到大数据对新闻行业的影响，至少体现在两个方面，一是应用技术来收集和深度分析数据；二是以可视化和互动式的效果呈现新闻。

3. 数据新闻实践：ProPublica 和 The Upshot

在美国，一提到数据新闻，人们最先想到的便是 ProPublica 和《纽约时报》推出的数据新闻专栏“The Upshot”。

总部设在美国纽约市曼哈顿区的非盈利公司 ProPublica 由《华尔街日报》前执行总编约翰·施泰格（Paul Steiger）创立，是一家独立的新闻编辑部，为公众利益进行调查报道。ProPublica 成立之初，便集合了一批最优秀的新闻人才。除了主编施泰格，还有来自《纽约时报》前调查报道主编斯蒂芬·恩格尔伯格（Stephen Engelberg）出任执行主编，记者和编辑全都来自传统媒体。

2011 年 4 月 18 日，新闻界最高荣誉、第 95 届普利策奖揭晓，ProPublica 以其作品《华尔街金钱机器》摘得国内新闻报道奖（2011 Pulitzer Prize for National Reporting）。该作品历时一年时间，通过可视化手段详尽地披露了华尔街金融巨头们为自己牟取巨额利润的全过程。

事实上，这已不是 ProPublica 第一次获得普利策奖。在 2010 年，也就是成立后的两年，它就凭借调查性报道与《纽约时报》共同获得了调查性报道奖，获奖记者谢里·芬克（Sheri Fink）写出了一篇关于奥尔良州一家医院在卡特里娜飓风中向病人注射致命毒剂的调查报道。该作品除了在 ProPublica 网站上登出之外，同样也在《纽约时报》上以纸质版面貌呈现，但像 2011 年完全以数字化报道摘得普利策奖，ProPublica 还属史上首次，由此它也成为美国数据新闻用户最活跃的网站之一。如今，ProPublica 的合作伙伴已有 104 家，如《纽约时报》《华盛顿邮报》《洛杉矶时报》《今日美国》MSNBC、CNN 等。

在 ProPublica 的记者看来，一个新闻应用就是一个可以讲故事的交互式数据库，你可以像对待任何一条新闻作品一样看待它，只不过是用软件替代了文字和图片。

“金钱医生”（Dollars for Docs）是 ProPublica 众多新闻应用中的一个。它追踪的是医药公司以咨询、讲课、旅游等形式付给医生数以亿计的额外费用。通过创建这一应用，读者能够查询到自己的医生，看他们收取酬劳的情况，其他合作新闻机构的记者也可以使用这些数据。

该应用问世后，总共有超过 125 家地方新闻机构的记者基于此数据对当地医生进行了调查报道，但这其中只有少数是 ProPublica 的正式合作伙伴，其余大多数都是使用这一新闻应用和数据进行独立报道。对此，ProPublica 团队认为，如果他们的数据能够帮助了解当地情况的报道者讲述有冲击力的故事，他们也就完成了自身的使命，这与 ProPublica 始终坚持的进行“真正重要”和“具有道德力量”的新闻报道的目标完全吻合。

如果说 ProPublica 是大数据时代的产物，那么 The Upshot 则是传统媒体进军数据新闻领域的典型。2014 年春天，《纽约时报》推出了新栏目 The Upshot，主打数据新闻。

该栏目由华盛顿分社前社长大卫·莱昂哈特（David Leonhardt）负责，成立之初便有 15 名团队成员，其中 3 位是全职图形设计。The Upshot 针对政治经济领域，旨在通过数据分析和呈现，帮助受众获知新闻背后的意义。在关注新闻之外，The Upshot 还设计了许多有趣的互动产品，通过数据分析和建模，受众可以按图索骥，找到自己需要的信息，并得到相关建议。

通过 The Upshot 和 ProPublica 的案例可以看出，数据新闻往往有三个特点。

其一，个性化。通过一组数据新闻或一款互动产品，受众能够产生共鸣，并找到与自己相匹配的答案。

其二，可视化。与传统新闻“文字 + 图片”不

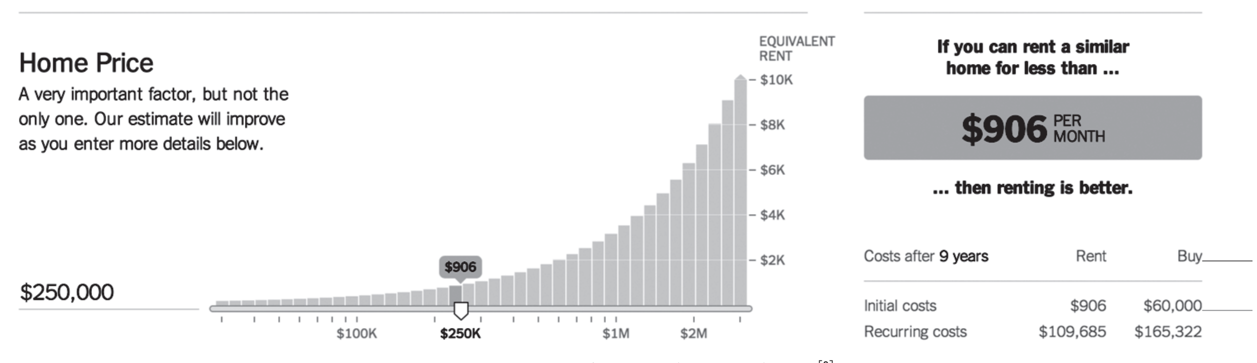


图 3 The Upshot 互动产品“租房还是买房？”[9]
（通过输入房屋价格、收入、居住时长、抵押情况、维护费用、税费、手续费和其他费用来测算是买房还是租房更合适。）

同，数据新闻更多的是用图表来讲故事，通过图形化的呈现和数据分析展现文字所无法达到的效果。如 The Upshot“美国生活最困难的地方是哪里？”，通过地图和一系列信息（整体排名、平均收入、教育程度、失业率、平均寿命等）展示每个地区的生活情况。这些反映政治、经济的主题看似宏观，但都与百姓生活密切相关，因而得到了受众的喜爱与好评。

其三，团队协作。在传统新闻时代，记者可以独立完成新闻，其与编辑、美编、校对是生产链的上下游。采编人员“生产”稿件，交由美编配图、排版，再通过印厂印刷完成。在数据新闻时代，新闻产品是团队合作的成果，在生产过程中，需要采编人员、设计人员和程序员共同制定方案，采编人员提供素材，程序员分析并处理数据，设计人员呈现可视化效果，三方是合作和互动的关系。媒

参考文献

[1] 黄志敏，王敏，李薇 . 数据新闻教育调查报告 . 新闻与写作，2017（9）.

[2] Jonathan Gray，Liliana Bounegru，Lucy Chambers.The Data Journalism Handbook.http：//datajournalismhandbook.org/1.0/en/，2012.

[3] Troy Thibodeaux.5 tips for getting started In data Journalism.http：//www.poynter.org/how-tos/digital-strategies/147734/5-tips-for-getting-started-in-data-journalism/，2011.

[4] 章戈浩 . 为开放新闻的数据新闻——英国《卫报》的数据新闻实践 . 新闻记者，2013（6）.

[5] 方洁，颜冬 . 作全球视野下的“数据新闻”：理念与实践 . 国际新闻界，2013（6）.

[6] Simon Rogers.Open data Journalism. http://simonrogers.net/2013/01/24/open-data-journalism/.

[7] Simon Rogers.A data journalism workflow. http://simonrogers.net/2013/01/27/a-data-journalism-workflow/.

[8] 图片来源：Brandshaw，P. The Inverted Pyramid of Data Journalism.http：//onlinejournalismblog.com/2011/07/07/the-inverted-pyramid-of-data-journalism/.

[9] 图片来源：http：//www.nytimes.com/interactive/2014/upshot/buy-rent-calculator.html.

（作者单位：《中国民航报社》）